Dual Dense Uncertainty Embedding for Iris Recognition

Yunlong Wang
MAIS, CASIA
yunlong.wang@ia.ac.cn

Jianze Wei
IME, CAS
weijianze@ime.ac.cn

Zhaofeng He
BUPT
zhaofenghe@bupt.edu.cn

Zhenan Sun
MAIS, CASIA
znsun@nlpr.ia.ac.cn

Tieniu Tan
CRIPAC, CASIA
tnt@nlpr.ia.ac.cn

Abstract

Pixelwise dense representations are more prevalent in the field of iris recognition, a.k.a. iris templates or IrisCodes. Almost all previous works of this kind are deterministic. To be specific, pixel-level representations are exclusively derived from certain point-by-point modeling, including filter response, phase correlations, ordinal relations and so on. Moreover, the binary mask indicating valid iris regions is solely determined by a fixed threshold or the output of standalone segmentation and localization algorithms. Uncertainty acquisition factors in the process of iris imagery formation are not considered. In this paper, we propose a simple yet effective building block termed as dual dense uncertainty embedding (D^2UE), which can be seamlessly incorporated into any deep learning (DL) frameworks that extract dense representations for iris recognition. D^2UE has two pathways wherein both take dense feature maps of the backbone network as input. One pathway of D^2UE predicts a variance-scaling map (VSM) and then applies it to an adaptive threshold masking operation on the iris image. The dynamic threshold for each pixel in this manner is dependent on not only the intensity distribution of the iris image but also each pixel's low-level uncertainty. The other pathway of D^2UE adopts over-parameterization technique and extracts uncertainty embedded dense representation (UEDR) by modeling each pixel's contextual uncertainty. Extensive experiments on several iris datasets demonstrate that recognition performance under both within-database and cross-database settings can be significantly improved by incorporating D^2UE into the baseline method. D^2UE only cost a slight computational overhead, but even surpassing a few SOTA methods with a large backbone network and much more training budget.

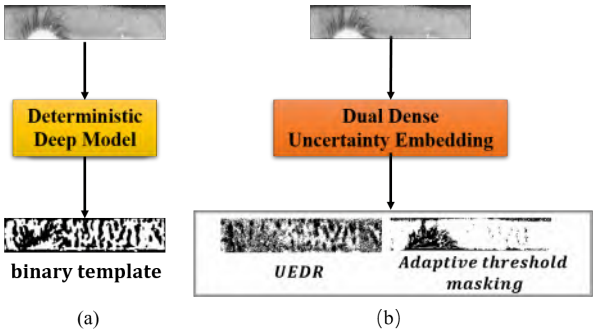


Figure 1. (a) Deep deterministic model outputs a binary iris template by certain point-by-point modeling, without considering uncertainty acquisition factors. In addition, the mask indicating valid iris region is provided by standalone iris segmentation methods. (b) The proposed D^2UE simultaneously extracts uncertainty embedded dense representation (UEDR) and yields the iris mask by adaptive threshold masking.

1. Introduction

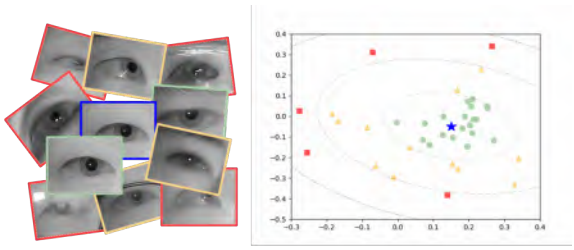
The iris of human eye is the annular, highly textured membrane that surrounds the pupil. Iris texture is unique and stable, so it serves as a desirable biometric trait. An iris recognition system exploits the richness of these textural patterns for human identification applications. In the early 1990s, a seminal work by John Daugman [4] introduced 2D Gabor filters to encode normalized iris images into a binary sequence of 2,048 bits, *a.k.a.* IrisCodes. Binary iris codes were achieved via phase demodulation and quantization. The Hamming distance was then used to evaluate the dissimilarity between the binary templates. Although numerous solutions came out over the last two decades, the series of approaches extracting pixel-level representations is still dominant in the field of iris image encoding and matching, *e.g.*, 1D Gabor filter response [9], phase correlations [12], ordinal relations [21] and so on.

In the era of deep learning, researchers begin to explore

the possibility of introducing deep dense representation into conventional iris recognition. UniNet [31] is a representative framework which adopts fully convolutional network (FCN) to generate spatially corresponding iris feature descriptors that obtains higher accuracy and generalization ability. [25] extended the framework of UniNet and proposed DRFNet with residual learning and dilated kernels. More literature work can be reviewed in a recent survey on deep learning for iris recognition [15]. Although these data-driven methods have achieved promising results, they have unconsciously stuck in feature ambiguity dilemma [20]. More specifically, the accuracy gain is marginal even if these models are trained on bigger iris datasets or adopt larger backbone network structure. As shown in Fig.2, the deterministic center represents an ideal identity feature that only contains personal identifiable information, while the deterministic points around the center are the iris features of the imagery affected by uncertain factors.

To our knowledge, most iris recognition methods map each iris sample to a deterministic point in latent feature space. However, uncertain acquisition factors inevitably affect the process of iris imagery formation. These factors can be categorized into three aspects, *i.e.* eye state (eyeball motion, squinting, gaze, occlusions), device configuration (optics, sensor, image signal processing) and environment (indoor/outdoor, ambient illumination). Training with these scattering deterministic points usually makes the model overfit non-identity information. [26] recently propose a DL framework that explores uncertain acquisition factors for more discriminative and robust iris recognition. Deep uncertainty embedding is leveraged to represent the iris image using a Gaussian distribution, wherein the mean encodes the most likely identity feature of the iris image, and the variance encodes the data uncertainty from acquisition factors. They desert the routine of dense iris representation but encode the data uncertainty along each dimension of a fixed-length feature vector. In addition, binary masks indicating valid iris regions are just neglected in their uncertainty embedding framework. In this paper, we integrate uncertainty learning into pixelwise dense iris recognition for the first time and revive the utilization of binary iris templates and masks. To sum up, the contributions of this paper are as follows:

- A simple yet effective building block termed as dual dense uncertainty embedding (D^2UE) is proposed. D^2UE can be attached to any DL frameworks that extract pixelwise dense iris representation.
- One pathway of D^2UE extracts uncertainty embedded dense representation (UEDR) in a pixel-dependent manner, while the other pathway predicts a variance-scaling map (VSM) and then embeds it to an adaptive thresholding mask operation.



(a) (b)

Figure 2. (a) Some example images from the same eye exhibit various uncertain factors in real-world iris acquisition. (b) Deterministic iris features extracted by UniNet and projected onto two-dimensional Cartesian space. Corresponding colors indicate the uncertainty (**variance**) of iris samples deviated from the identity (**mean**).

- Experimental results verify that the performance gain by incorporating D^2UE into the baseline method is encouraging, even surpassing a few SOTA methods with a large backbone network. In addition, the computational cost of D^2UE is slight.

2. Related Work

Deep Iris Feature Extraction Automatic feature learning is acknowledged as an overwhelming advantage of DL-based methods over traditional handcrafted features. The first iris recognition method using DL [10] is proposed for spoofing detection in 2015. DeepIris proposed by [8] establishes the relationship between heterogeneous iris images by learning a pairwise filter bank. DeepIrisNet [5] utilizes deep convolutional neural networks (CNNs) as feature extractors and applies to cross-sensor iris recognition. UNiNet [31] has two separate FCN-style models called *FeatNet* and *MaskNet* for iris feature and mask extraction respectively. [30] proposes a novel network architecture with Maxout units as the activation function and jointly optimizes the representation of iris and periocular biometrics. [25] extends the framework of UniNet and proposes DRFNet with residual learning and dilated kernels. [27] narrows the distribution gap of cross-sensor iris datasets by utilizing sensor-specific information and adversarial strategy. [18] combines the merits of both CNNs and graph models to learn dynamic graph representations for occluded iris recognition. The idea of network architecture search (NAS) is first introduced in [13] to search the resource-constrained network architecture for iris recognition. [14] proposed a fully complex-valued neural network, which captures the multi-scale, multi-resolution, and multi-orientation phase and amplitude features of iris texture. All of these methods map each iris sample to a deterministic point in the form of binary template or feature vector. They strive to narrow the intra-class distances and enlarge the inter-class distances for

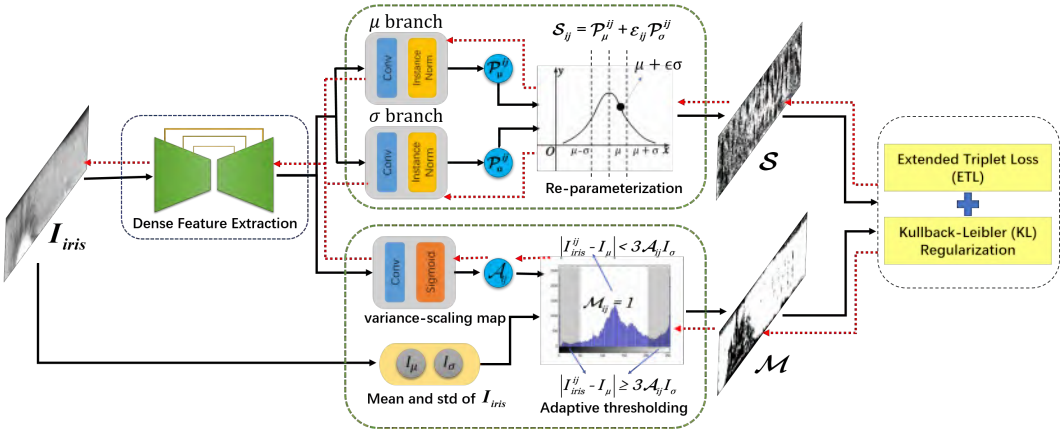


Figure 3. The schematic of the proposed framework D^{2}UE.

performance improvement but neglect the uncertain factors in iris acquisition.

Uncertainty Learning Deep uncertainty learning (DUL) is becoming a hotspot over the past several years. Approaches related to DUL can be roughly classified into two categories, i.e., data uncertainty [24] and model uncertainty [17]. Recently, DUL methods have made impressive progress in a few vision tasks. [6] proposes Bayesian SegNet for semantic segmentation, which utilizes Monte Carlo sampling with dropout to measure model uncertainty for each class. [3] proposes Gaussian YOLOv3 for object detection, wherein the loss function was redesigned by introducing the Gaussian function and the reliability of the bounding box was indicated by the localization uncertainty. For biometric recognition tasks, several work [20, 7, 2] have investigated the uncertainty modeling in face recognition. [20] discusses feature ambiguity dilemma with deterministic face embeddings and propose probabilistic face embeddings (PFEs). [7] adopts the Bayesian uncertainty estimates to measure the difficulty level of individual samples and refine the classification boundaries. [2] simultaneously learns a Gaussian distributional estimation for each face image with both identity feature and uncertainty. Similarly, [29] addresses both the noisy label and outlying sample problems in ReID tasks by explicitly modeling feature uncertainty as a Gaussian distribution.

More recently, [26] proposes the first DL work that leverages DUL for iris recognition. Specifically, uncertainty embedding (UE) and uncertainty-guided curriculum learning (UGCL) mitigate the influence of uncertain factors during iris acquisition. Note that [26] is most relevant to our work. But the prominent differences are that [26] represents iris images as compact feature vectors rather than dense representations, *i.e.*, binary iris templates and masks in our work. [26] directly extracts uncertainty embedding from the normalized iris images and does not use any binary masks indicating valid iris regions. In addition, bit-shifting operation for eye rotation mitigation is also neglected.

3. Framework

The schematic of the proposed D^{2}UE is depicted in Fig.3. The details will be discussed in this section.

3.1. Preliminary

Deterministic Dense Representation. A large portion of current iris recognition methods represent each iris sample as a deterministic binary template. An iris image is usually preprocessed and normalized to a rectangular image I_{iris} with the resolution $H \times W$. I_{iris} is fed to a backbone sub-network $F_\theta(\cdot)$ with a group of parameters θ. Without losing spatial resolution, it yields dense feature maps $\mathcal{H} \in \mathbb{R}^{C\times H\times W}$ as Eq.1.

$$\mathcal{H} = F_\theta(I_{iris}) \tag{1}$$

where C denotes the channel number of $\mathcal{H}$. $\mathcal{H}$ is then passed to a feature extraction sub-network $F_\varphi(\cdot)$ with a group of parameters φ. A single-channel feature map $\mathcal{P} \in \mathbb{R}^{1\times H\times W}$ is obtained as Eq.2.

$$\mathcal{P} = F_\varphi(\mathcal{H}) \tag{2}$$

To generate the final representation $\mathcal{T} \in \{0,1\}^{1\times H\times W}$ in the form of binary template, $\mathcal{P}$ is binarized via some certain thresholding techniques, *e.g.*, a fixed value η as Eq.3.

$$\mathcal{T}_{ij} = \begin{cases} 1, & if\ \mathcal{P}_{ij} > \eta \\ 0, & otherwise \end{cases} \tag{3}$$

These deterministic methods for iris recognition strive to narrow the intra-class distances and enlarge the inter-class distances of iris samples through data-driven optimization, but the uncertainties in iris acquisition hinder them from learning more discriminative representations. In this paper, we incorporate uncertainty modeling into dense representation learning to solve this bottleneck problem.

Deterministic Threshold Masking Strategy. OSIris is a popular open-source iris recognition software [16]. It leverages the characteristics of iris intensity distribution and provides a deterministic threshold masking way to mask out noises on normalized iris images as Eq.4. I_μ and I_σ are the mean and standard deviation of the normalized iris image I_{iris}. i and j are the row and column indices of one pixel. I_{mask} is a binary mask indicating where valid iris pixels are, *i.e.*, pixels belonging to visible iris texture are set to 1 while noisy pixels are set to 0.

$$I^{ij}_{mask} = \begin{cases} 1, & if \ \left| I^{ij}_{iris} - I_\mu \right| < \alpha I_\sigma \\ 0, & otherwise \end{cases} \tag{4}$$

The ground of this strategy lies in that the intensity distribution of valid iris pixels is inside $(I_\mu - \alpha I_\sigma, I_\mu + \alpha I_\sigma)$. It is generally based on the observation that the noises including specular reflections and eyelids are usually bright, and noisy pixels of eyebrows and pupil are in low gray values. Hence, α is a key hyperparameter which controls the allowable intensity range of valid iris pixels. Obviously, a larger α means more pixels will be absorbed as iris textures but tends to mislabel non-iris pixels.

In the implementation of OSIris [16], α is empirically fixed to 2.35. It is verified on several earlier small-scale iris datasets to separate iris pixels and non-iris pixels. However, a deterministic α is suboptimal considering the versatile uncertainties in the intensity distribution of iris images collected from various scenarios.

3.2. Uncertainty Embedded Dense Representation

Uncertainty embedding is for the first time introduced to represent an iris image using a multivariate Gaussian distribution rather than a deterministic point in [26]. A two-branch structure encodes the mean and variance of the virtual distribution respectively. Next, the re-parameterization technique [11] is adopted to generate virtual features by sampling random noise from a normal distribution. Different from the way that [26] constructs uncertainty embedded feature vectors, we implement the proposed UEDR in a pixel-dependent manner.

Specifically, to model each pixel's uncertainty by a univariate Gaussian distribution, μ-branch and σ-branch are constructed by a simple transformation with a convolutional layer and an instance normalization (IN) layer [22]. μ-branch takes $\mathcal{H}$ as input and outputs $\mathcal{P}_\mu \in \mathbb{R}^{1\times H\times W}$ as Eq.5, where $Conv(\cdot)$ denotes a convolutional layer and $IN(\cdot)$ denotes an IN layer.

$$\mathcal{P}_\mu = IN_\mu(Conv_\mu(\mathcal{H})) \tag{5}$$

Similarly, σ-branch yields $\mathcal{P}_\sigma \in \mathbb{R}^{1\times H\times W}$ as Eq.6.

$$\mathcal{P}_\sigma = IN_\sigma(Conv_\sigma(\mathcal{H})) \tag{6}$$

In contrast to [26] which encodes the data uncertainty along each dimension of a fixed-length feature vector, a pixel-relevant standard normal distribution $\varepsilon_{ij} \sim \mathcal{N}(0,1)$ is leveraged to sample the random noise and generate the virtual pixel-wise representation as Eq.7.

$$\mathcal{S}_{ij} = \mathcal{P}^{ij}_\mu + \varepsilon_{ij} \times \mathcal{P}^{ij}_\sigma \tag{7}$$

$\mathcal{S} \in \mathbb{R}^{1\times H\times W}$ is the final UEDR of the input iris image. During training, $\mathcal{S}$ lets the gradient backpropagate through the re-parameterization operation and optimize the objective function. But in inference phase, $\mathcal{P}_\sigma$ which encodes uncertainty, is not helpful for the improvement of recognition performance. Hence, σ-branch is abandoned and only μ-branch is retained to compute iris representations for feature matching during inference as Eq.8.

$$\mathcal{S}_{ij} = \mathcal{P}^{ij}_\mu \tag{8}$$

3.3. Adaptive Threshold Masking

As mentioned above, the deterministic threshold masking strategy as Eq.4 in OSIris [16] can not adapt to various data uncertainties, *e.g.*, illumination variations and noisy outliers. In this paper, we leverage the power of DUL to find a optimal α for each pixel in an iris image, taking not only the intensity distribution of the iris image but also each pixel's low-level uncertainty into consideration. Specifically, a variance scaling map (VSM), *i.e.*, $\mathcal{A} \in (0,1)^{1\times H\times W}$ is obtained via Eq.9. To be concrete, the dense feature map $\mathcal{H}$ is passed to a convolutional layer which outputs a single-channel pixelwise logit map and then activated by Sigmoid function.

$$\mathcal{A} = Sigmoid(Conv(\mathcal{H})) \tag{9}$$

In contrast to a fixed α for all the pixels in an iris image, 3σ rule is adopted here and each pixel's adaptive threshold is both dependent on $\mathcal{A}_{ij}$, I_μ and I_σ as Eq.10.

$$\mathcal{M}_{ij} = \begin{cases} 1, & if \ \left| I^{ij}_{iris} - I_\mu \right| < 3\mathcal{A}_{ij} I_\sigma \\ 0, & otherwise \end{cases} \tag{10}$$

It can be seen from Eq.10 that a pixel is deemed as belonging to valid iris region if its intensity lies in $(I_\mu - 3\mathcal{A}_{ij} I_\sigma, I_\mu + 3\mathcal{A}_{ij} I_\sigma)$. In this manner, the adaptive threshold masking here will treat each pixel differently according to its low-level uncertainty conveyed by $\mathcal{A}_{ij}$ in an iris image. For instance, a partially over-illuminated iris image will drive some bright pixels' $\mathcal{A}_{ij}$ to be larger than other pixels inside normally-illuminated regions.

Note that there are no ground truth binary mask here as supervision signal to optimize the group of parameters to obtain $\mathcal{A}$. This branch for adaptive threshold masking is also supervised by the iris images' ID labels as the UEDR

branch and jointly optimized in an end-to-end training manner. More details about network training will be provided in the following subsection.

3.4. Loss Function

Extended Triplet Loss (ETL) Triplet loss [19] is widely used in metric learning based methods. A variant of Triplet Loss is employed, dubbed as Extended Triplet Loss (ETL) in UniNet [31] that is tailored for iris templates as Eq.11.

$$\mathcal{L}_{etl} = \frac{1}{M}\sum_{i=1}^{M}\max(D(\mathcal{S}_i^P, \mathcal{S}_i^A) - D(\mathcal{S}_i^N, \mathcal{S}_i^A) + \tau, 0) \tag{11}$$

where M is the number of triplets in a mini-batch, τ is a preset parameter to control the desired margin between anchor-positive and anchor-negative distance. A triplet here consists of UEDRs and binary iris masks from anchor, positive and negative iris images respectively, *i.e.*, $\mathcal{S}_i^A, \mathcal{S}_i^P, \mathcal{S}_i^N$ and $\mathcal{M}_i^A, \mathcal{M}_i^P, \mathcal{M}_i^N$. Different from Euclidean distance or cosine similarity adopted in deep face recognition, the computation of the distance between the UEDRs of two iris images is more complicated. It is mainly because non-iris region masking and horizontal bit shifting to mitigate eye rotation should be taken into consideration. The minimum shifted and masked distance (MMSD) is adopted as Eq.12 to compute the distance between two UEDRs and corresponding binary masks.

$$D(\mathcal{S}^1, \mathcal{S}^2) = \min_{-B\leq b\leq B}\left\{FD(\mathcal{S}_b^1, \mathcal{S}^2)\right\} \tag{12}$$

where B is the maximum pixel that bit-shifting would horizontally shift the UEDR. FD is fractional distance. It is used for measuring the distances at valid iris positions where both $\mathcal{M}_{ij}^1$ and $\mathcal{M}_{ij}^2$ equal to 1, and normalizing the total distance by the number of valid iris pixels $|m|$ as Eq.13.

$$\begin{aligned} FD(\mathcal{S}^1, \mathcal{S}^2) &= \tfrac{1}{|m|}\sum_{(i,j)\in m}\left(S_{ij}^1 - S_{ij}^2\right)^2 \\ m &= \{(i,j) \mid \mathcal{M}_{ij}^1 = 1 \; and \; \mathcal{M}_{ij}^2 = 1\} \end{aligned} \tag{13}$$

Kullback-Leibler (KL) Regularization If only training with ETL loss, the model will tend to predict UEDR with smaller variance σ which significantly weakens its ability to modeling uncertainty. Hence, a Kullback-Leibler (KL) regularization term is adopted for the constraint on each pixel's variance in UEDR. The KL regularization term forces each pixel's predicted distribution $\mathcal{N}(\mathcal{S}_{ij}|\mu_{ij}, \sigma_{ij})$ to be close to a normal Gaussian distribution $\mathcal{N}(\varepsilon_{ij}|0, 1)$ as Eq.14.

$$\begin{aligned} \mathcal{L}_{kl} &= \mathrm{E}\left\{KL\left[\mathcal{N}(\mathcal{S}_{ij}|\mu_{ij}, \sigma_{ij}) \parallel \mathcal{N}(\varepsilon_{ij}|0, 1)\right]\right\} \\ &= -\tfrac{1}{2NHW}\sum_{n=1}^{N}\sum_{h=1}^{H}\sum_{w=1}^{W}(1 + \log\sigma_{n,i,j}^2 - \mu_{n,i,j}^2 - \sigma_{n,i,j}^2) \end{aligned} \tag{14}$$

The final total loss for model training is

$$\mathcal{L}_{total} = \mathcal{L}_{etl} + \lambda\mathcal{L}_{kl} \tag{15}$$

where λ is the trade-off parameter.

Ent-to-end Training As indicated by the red dotted lines in Fig.3, there are no additional supervisions except $\mathcal{L}_{total}$ in the proposed D^{2}UE. The supervision signal is back propagated to both UEDR and VSM branches for parameter updating. In a nutshell, D^{2}UE can be trained in an end-to-end strategy. Most current dense iris representation methods require a segmented binary mask provided by standalone iris segmentation methods. But both iris template and mask are simultaneously obtained in D^{2}UE.

3.5. Implementation Details

Iris Image Preprocessing For fair comparisons, the same iris image preprocessing methods are adopted for the proposed D^{2}UE and the compared methods. Specifically, each iris image was first processed by IrisParseNet [23] to obtain the binary iris mask. Some falsely segmented iris samples are double-checked and corrected manually. The Hough circle detection [28] was used to fit the centers and the radii of iris and pupil. Next, circular iris region of each iris image was normalized to a 64×512 rectangular image by the rubber-sheet model [4]. We apply the same contrast enhancement process as UniNet [31] to the normalized iris images. The enhanced images are used as input for model training and test.

Hyperparameter Setting The kernel size of convolutional layers in generating UEDR and VSM is all set to 3. The proposed framework is implemented using PyTorch. The optimizer is Adam for parameter updating. The batch size is set to 64, *i.e.*, 64 triplets in a mini-batch. The marginal distance τ in ETL loss is set to 0.2. The trade-off parameter λ to balance ETL loss and KL regularization is set to 0.01. The learning rate (LR) is initially set to $1e-3$ and a cosine annealing schedule is adopted to adjust LR in each training epoch.

Training and Inference Details The proposed D^{2}UE was trained around 50 to 100 epochs until convergence and each epoch took a dozen of minutes on a GPU server equipped with one Nvidia GeForce RTX 3090 card. During training, UEDRs of iris triplet samples contain real values, and it is more convenient to compute the fractional Euclidean distance as Eq.13. As mentioned earlier, binary iris templates or IrisCodes are more lightweight and efficient. A similar feature binarization step is adopted as UniNet [31]

Table 1. The detailed information of adopted iris datasets.

Dataset	Training Set		Test Set	
	#Class	#Images	#Genuine	#Imposter
ND0405	356	7,925	14,791	5,743,130
Distance	142	1,799	20,702	2,969,533
Thousand	1,000	10,000	45,000	49,950,000
Mobile	180	1,800	8,100	1,611,000

which takes the mean value as the threshold to binarize the output iris representation in Eq.8 during inference. For feature matching, we use fractional Hamming distance as the metric to compute the dissimilarity from the binarized iris representation and the binary mask predicted by adaptive threshold masking.

4. Experiments

4.1. Databases and Protocols

The following four iris datasets are employed to verify the efficacy of the proposed framework D^{2}UE. The detailed information of these datasets are tabulated in Table 1.

ND-IRIS-0405 Iris Image Dataset This database [1] consists of 64980 iris images from 356 subjects. The dataset splitting protocol is the same as UniNet [31].

CASIA Iris Image Dataset V4-Distance This database is released by Chinese Academy of Sciences, Institute of Automation (CASIA). It has 2567 iris samples from 142 subjects. The dataset splitting protocol is the same as UniNet [31].

CASIA Iris Image Dataset V4-Thousand This database contains 20,000 iris images from both eyes of 1,000 objects. The same dataset splitting protocol is adopted as [26].

CASIA Iris Image Dataset Mobile CASIA built this dataset for mobile iris recognition using mobile phones. The same dataset splitting protocol is adopted as [26].

4.2. Compared Methods and Metrics

As a building block, the proposed D^{2}UE can be incorporated into any DL frameworks that extract dense representations for iris recognition. In this paper, UniNet [31] is taken as baseline and use the stack of feature maps fed to $Conv4$ in *FeatNet* as the input of two pathways, *i.e.*, UEDR and adaptive threshold masking. Several representative non-DL based and DL based methods are compared with D^{2}UE, including log-Gabor coding [4], ordinal measure (OM) coding [21], Maxout [30], the baseline UniNet [31]. In addition, we also compare with the first uncertainty embedding method for iris recognition [26]. For log-Gabor coding [4], the wavelength is set to 18 and the standard deviation to 0.5$\times$wavelength. The bit-shift operation shifts the pixel in the range of [-8, 8]. For OM coding, the two-lobe model with an inter-lobe distance of 7 and orientation of 0 is selected. We retrain the released models of Maxout [30], UniNet [31] and [26] under the same dataset splitting protocols.

The detection error trade-off (DET) curve is adopted as the evaluation metrics. It gives all possible operating points, and is typically depicted as false non-match rate (FNMR) versus false match rate (FMR). FNMRs are compared at FMR = 0.1%, and equal error rate (EER), i.e., FNMR=FMR is also utilized for evaluating recognition performances of the compared methods.

4.3. Within-database Comparisons

In the within-database settings, the algorithms are trained and tested under the same data splitting protocol on one iris database. The results of within-database comparisons are tabulated in Tab.2. It can be obviously seen that the proposed D^{2}UE improves the recognition performance of the baseline method UniNet by a large margin on all the iris databases. It demonstrates that D^{2}UE can extract more discriminative dense iris representations, mainly due to modeling pixel-level contextual uncertainty. Although D^{2}UE is a bit inferior to the recent SOTA method [26] in some comparisons, D^{2}UE gains huge advantages over [26] in small model size and better computational efficiency which will be discussed in the following subsection.

4.4. Cross-database Comparisons

In the cross-database settings, the algorithms are trained on ND-Iris-0405 and tested on CASIA-IrisV4-Distance, CASIA-IrisV4-Thousand and CASIA-Iris-Mobile respectively. It can investigate the generalization ability of the compared methods when applied in real-world scenarios. The results of cross-database comparisons are listed in Tab.3. The proposed D^{2}UE performs rather well on unseen iris data. It achieves better performance than the baseline method UniNet on two of the three iris databases and comparable on CASIA-IrisV4-Thousand.

4.5. Ablation Study and Visualizations

To verify the effectiveness of UEDR and adaptive thresholding mask in D^{2}UE, ablation study is conducted on CASIA-IrisV4-Distance under the within-database protocol. We also compute the model complexity, including the number of parameters and multiply–accumulate operations (MACs), to demonstrate the efficacy of D^{2}UE.

Effectiveness of UEDR. In addition to UEDR, two ablation models are configured and compared, *i.e.*, *FeatNet* and *FeatNet+*. *FeatNet* is of the same network structure and configuration as UniNet in [31]. Considering that UEDR has two branches and each occupies one convolutional layer, one convolutional layer with the kernel $N \times 2 \times 3 \times 3$ is inserted after $Conv4$ in *FeatNet* and adjust the kernel size

Table 2. Within-database comparisons measured by FNMR(%) at **0.1%** FMR and EER(%). **Bold** texts indicate the best and underlined texts indicate the second best. "↓" denotes the reduction of error rate on the baseline method that has a superscript symbol "†".

		logGobar	Ordinal	Maxout	Wei et al.	UniNet†	Ours
ND0405	EER	1.94	2.49	7.33	5.05	1.74	**1.61** (↓0.13)
	FNMR	3.24	7.51	32.31	20.90	2.75	**2.11** (↓0.64)
Distance	EER	2.60	8.01	3.73	**1.17**	2.17	1.57 (↓0.60)
	FNMR	6.61	40.73	14.85	4.18	5.32	**3.95** (↓1.37)
Thousand	EER	1.97	1.75	3.78	1.17	1.22	**0.85** (↓0.37)
	FNMR	4.81	4.62	16.24	4.17	2.73	**1.51** (↓1.22)
Mobile	EER	9.24	3.62	2.85	**0.79**	2.11	1.56 (↓0.55)
	FNMR	37.94	19.73	10.58	**2.67**	12.50	4.25 (↓8.25)

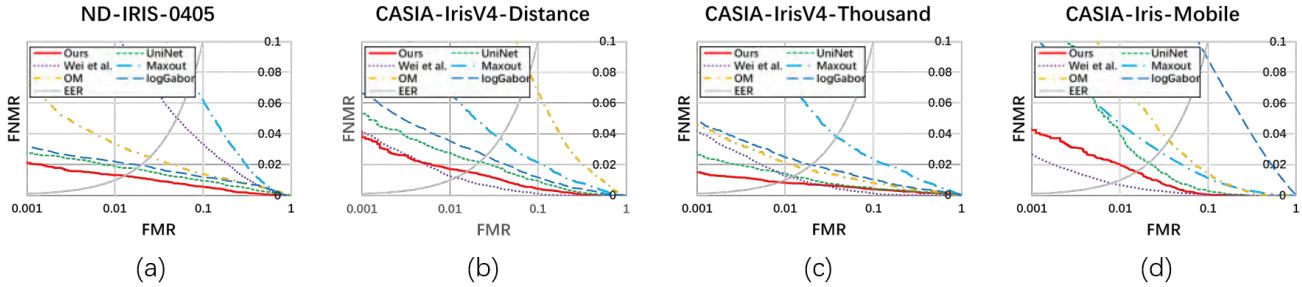


Figure 4. DET curves of within-database comparisons on (a) ND0405, (b) CASIA-IrisV4-Distance, (c) CASIA-IrisV4-Thousand, (d) CASIA-Iris-Mobile.

Table 3. Cross-database comparisons measured by FNMR(%) at **0.1%** FMR and EER(%).

	UniNet		Ours	
	EER	FNMR	EER	FNMR
Distance	7.55	23.48	7.04	23.21
Thousand	3.10	10.68	5.3	13.82
Mobile	1.31	4.87	1.18	2.79

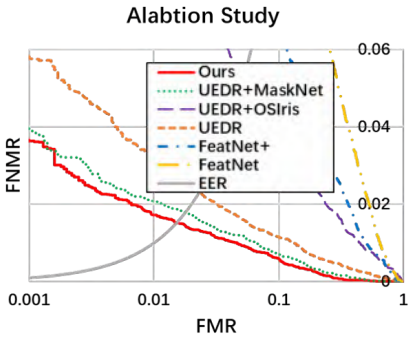


Figure 5. DET curves of ablation study.

of the output layer to $2 \times 1 \times 3 \times 3$. N is the channel number of the dense feature map output by $Conv4$. This variant of *FeatNet* is dubbed as *FeatNet+*. *FeatNet+* has nearly an equal number of network parameters as UEDR (actually 18 more). Note that no iris masks are utilized here.

As shown in Fig.5 and tabulated in Tab.4, UEDR significantly outperforms both *FeatNet* and *FeatNet+*. The improvement on recognition performance is large, *i.e.*, EER reduced from 11.97% to 2.27%. Laterally, the performance gain over *FeatNet+* indicates that UEDR is much more powerful than simply adding more layers in the baseline network.

Effectiveness of Adaptive Threshold Masking. Two methods to generate binary iris masks, *i.e.*, *OSIris* and *MaskNet*, are taken as ablation models. We adopt the officially released model of *MaskNet* in UniNet [31]. Note that UEDR only is regarded as the ablation model with no iris mask.

The effectiveness of adaptive threshold masking can be justified from Fig.5 and Tab.4. The EER of UEDR is reduced from 2.27% to 1.57% by joint adaptive threshold masking. The deterministic threshold masking of *OSIris* performances even worse than UEDR only. It implies that the masks obtained in this way may harm the recognition ability of UEDR and can not adapt to various uncertainties in iris acquisition.

Model Complexity. D^{2}UE models pixelwise uncertainty in quite a lightweight way. The model complexity is calculated using THOP[1] and listed in Tab.5. Compared with the recent SOTA methods [26] and the baseline [31], D^{2}UE is very tiny in model size and only cost a slight computational overhead. To be concrete, D^{2}UE has much fewer parameters and a bit less MACs compared to UniNet (*FeatNet+MaskNet*), but obtains better recognition performance.

[1]https://github.com/Lyken17/pytorch-OpCounter

Table 4. The configurations of ablation models and corresponding EERs (%).

	No Iris Mask	Deterministic Threshold Masking	Adaptive Threshold Masking	EER
FeatNet	✔			11.97
FeatNet+	✔			7.61
UEDR	✔			2.27
UEDR+*OSIris*		✔		5.25
UEDR+*MaskNet*		✔		1.79
Ours			✔	**1.57**

Table 5. The model complexity measured by #Params and MACs.

		#Params(K)	MACs(M)
Wei et al.		6558.11	1261.37
UniNet	*FeatNet*	27.46	144.90
	MaskNet	97.86	98.86
	Total	125.32	243.76
Ours		**29.48**	**211.22**

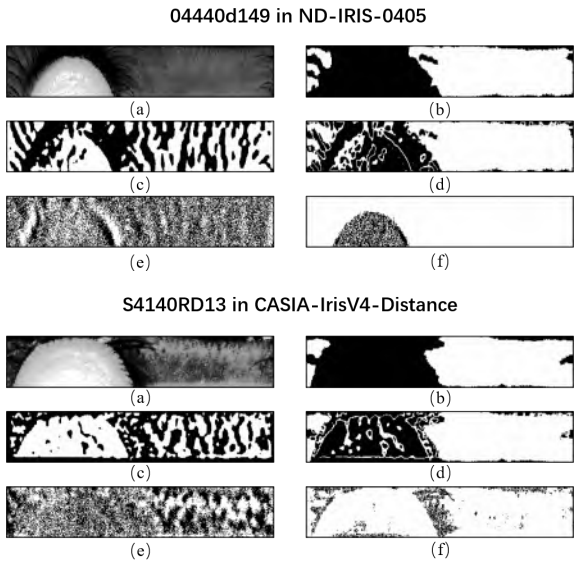


Figure 6. Visualizations of some exemplar UEDRs and binary masks predicted by adaptive thresholding mask. (a) Normalized iris image. (b) Manually labelled iris mask. (c) Binary iris template generated by *FeatNet* in UniNet. (d) Binary iris mask predicted by *MaskNet* in UniNet. (e) Our binary UEDR. (f) Our binary iris mask predicted by adaptive threshold masking.

Visualizations. Some exemplar UEDRs and binary masks predicted by adaptive thresholding masking are visualized in Fig.6, also with normalized iris image, manually labelled iris mask, binary iris templates and masks generated by UniNet. Although binary UEDRs have been squeezed and quantized for efficient computations, it can still be observed that the UEDRs effectively avoids some ineffective areas such as eyebrows and eyelids. More interestingly, it works in a way quite distinct from the way UniNet does. Uncertainty modeling in D^{2}UE results in coarse UEDRs and irregular iris masks. The underlying reasons may be that UEDR and prediction of variance-scaling map (VSM) are jointly trained in an end-to-end manner and supervised by only identity labels. Our D^{2}UE improves the recognition performance of the baseline methods by assembling UEDR and adaptive thresholding masking together. It also implies that it may be unnecessary to tediously annotate or segment each ocular pixel into semantic attributes. An implicit end-to-end learning way would probably be better, e.g., the proposed adaptive threshold masking.

5. Conclusion

In this paper, D^{2}UE is proposed to model data uncertainty of iris recognition in a pixel-level manner, which attenuates the interference caused by uncertain acquisition factors. D^{2}UE takes the intermediate feature map of any dense DL framework as input and generates the UEDR and VSM of an iris image. VSM is then leveraged to produce a binary mask through adaptive threshold masking, and thus pixelwise iris segmentation is on longer needed. Experimental results on several iris datasets demonstrate the superiority of D^{2}UE in improving the recognition performance of baseline methods, and it is a remarkably lightweight building blocking. We will extend D^{2}UE to other upcoming backbone networks and verify its generality.

6. Acknowledgement

This work is supported by the National Key Research and Development Program of China under Grant No. 2022YFC3310400, and the National Natural Science Foundation of China under Grant No. 62006225, 62276263, also sponsored by the Beijing Nova Program under Grant No. Z201100006820050, Z211100002121010.

References

[1] K. W. Bowyer and P. J. Flynn. The nd-iris-0405 iris image dataset. *Notre Dame CVRL Technical Report*, 2016.

[2] J. Chang, Z. Lan, C. Cheng, and Y. Wei. Data uncertainty learning in face recognition. In *CVPR*, pages 5710–5719, 2020.

[3] J. Choi, D. Chun, H. Kim, and H.-J. Lee. Gaussian yolov3: An accurate and fast object detector using localization uncertainty for autonomous driving. In *ICCV*, pages 502–511, 2019.

[4] J. G. Daugman. High confidence visual recognition of persons by a test of statistical independence. *IEEE TPAMI*, 15(11):1148–1161, 1993.

[5] A. Gangwar and A. Joshi. Deepirisnet: Deep iris representation with applications in iris recognition and cross-sensor iris recognition. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 2301–2305. IEEE, 2016.

[6] S. Isobe and S. Arai. Deep convolutional encoder-decoder network with model uncertainty for semantic segmentation. In *2017 IEEE International Conference on INnovations in Intelligent SysTems and Applications (INISTA)*, pages 365–370. IEEE, 2017.

[7] S. Khan, M. Hayat, S. W. Zamir, J. Shen, and L. Shao. Striking the right balance with uncertainty. In *CVPR*, pages 103–112, 2019.

[8] N. Liu, M. Zhang, H. Li, Z. Sun, and T. Tan. Deepiris: Learning pairwise filter bank for heterogeneous iris verification. *Pattern Recognition Letters*, 82:154–161, 2016.

[9] L. Masek et al. *Recognition of human iris patterns for biometric identification*. PhD thesis, Master's thesis, University of Western Australia, 2003.

[10] D. Menotti, G. Chiachia, A. Pinto, W. R. Schwartz, H. Pedrini, A. X. Falcao, and A. Rocha. Deep representations for iris, face, and fingerprint spoofing detection. *IEEE Transactions on Information Forensics and Security*, 10(4):864–879, 2015.

[11] A. C. Miller, N. J. Foti, and R. P. Adams. Variational boosting: Iteratively refining posterior approximations. In *International Conference on Machine Learning*, pages 2420–2429. PMLR, 2017.

[12] D. M. Monro, S. Rakshit, and D. Zhang. Dct-based iris recognition. *IEEE transactions on pattern analysis and machine intelligence*, 29(4):586–595, 2007.

[13] K. Nguyen, C. Fookes, and S. Sridharan. Constrained design of deep iris networks. *IEEE Transactions on Image Processing*, 29:7166–7175, 2020.

[14] K. Nguyen, C. Fookes, S. Sridharan, and A. Ross. Complex-valued iris recognition network. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–1, 2022.

[15] K. Nguyen, H. Proença, and F. Alonso-Fernandez. Deep learning for iris recognition: A survey. *arXiv preprint arXiv:2210.05866*, 2022.

[16] N. Othman, B. Dorizzi, and S. Garcia-Salicetti. Osiris: An open source iris recognition software. *Pattern Recognition Letters*, 82:124–131, 2016.

[17] J. Postels, F. Ferroni, H. Coskun, N. Navab, and F. Tombari. Sampling-free epistemic uncertainty estimation using approximated variance propagation. In *ICCV*, pages 2931–2940, 2019.

[18] M. Ren, Y. Wang, Z. Sun, and T. Tan. Dynamic graph representation for occlusion handling in biometrics. In *AAAI*, pages 11940–11947, 2020.

[19] F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. In *CVPR*, pages 815–823, 2015.

[20] Y. Shi and A. K. Jain. Probabilistic face embeddings. In *ICCV*, pages 6902–6911, 2019.

[21] Z. Sun and T. Tan. Ordinal measures for iris recognition. *IEEE Transactions on pattern analysis and machine intelligence*, 31(12):2211–2226, 2008.

[22] D. Ulyanov, A. Vedaldi, and V. Lempitsky. Instance normalization: The missing ingredient for fast stylization. *arXiv preprint arXiv:1607.08022*, 2016.

[23] C. Wang, J. Muhammad, Y. Wang, Z. He, and Z. Sun. Towards complete and accurate iris segmentation using deep multi-task attention network for non-cooperative iris recognition. *IEEE Transactions on information forensics and security*, 15:2944–2959, 2020.

[24] G. Wang, W. Li, M. Aertsen, J. Deprest, S. Ourselin, and T. Vercauteren. Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with convolutional neural networks. *Neurocomputing*, 338:34–45, 2019.

[25] K. Wang and A. Kumar. Toward more accurate iris recognition using dilated residual features. *IEEE Transactions on Information Forensics and Security*, 14(12):3233–3245, 2019.

[26] J. Wei, H. Huang, Y. Wang, R. He, and Z. Sun. Towards more discriminative and robust iris recognition by learning uncertain factors. *IEEE Transactions on Information Forensics and Security*, 17:865–879, 2022.

[27] J. Wei, Y. Wang, X. Wu, Z. He, R. He, and Z. Sun. Cross-sensor iris recognition using adversarial strategy and sensor-specific information. In *2019 IEEE 10th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–8. IEEE, 2019.

[28] R. P. Wildes. Iris recognition: an emerging biometric technology. *Proceedings of the IEEE*, 85(9):1348–1363, 1997.

[29] T. Yu, D. Li, Y. Yang, T. M. Hospedales, and T. Xiang. Robust person re-identification by modelling feature uncertainty. In *ICCV*, pages 552–561, 2019.

[30] Q. Zhang, H. Li, Z. Sun, and T. Tan. Deep feature fusion for iris and periocular biometrics on mobile devices. *IEEE Transactions on Information Forensics and Security*, 13(11):2897–2912, 2018.

[31] Z. Zhao and A. Kumar. Towards more accurate iris recognition using deeply learned spatially corresponding features. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3809–3818, 2017.